

RESEARCH ARTICLE

Preparation of multiplexed control materials for cancer mutation analysis by genome editing in GM12878 cells

Guigao Lin^{1,2,3}  | Kuo Zhang^{1,2,3} | Yanxi Han^{1,3} | Rongxue Peng^{1,3} | Jinming Li^{1,2,3} 

¹National Center for Clinical Laboratories, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Hospital/National Center of Gerontology, Beijing, China

²Graduate School, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

³Beijing Engineering Research Center of Laboratory Medicine, Beijing Hospital, Beijing, China

Correspondence

Jinming Li, National Center for Clinical Laboratories, Beijing Hospital, No.1 Dahua Road, Dongdan, Beijing 100730, China. Email: jmli@nccl.org.cn

Funding information

Beijing Municipal Natural Science Foundation, Grant/Award Number: 7192188 and 7204299; Beijing Dongcheng District Outstanding Talent Nurturing Program, Grant/Award Number: 2019WJGW-10-01; National Natural Science Foundation of China, Grant/Award Number: 81974319 and 81902145

Abstract

Background: Quality control materials are necessary for assay development, test validation, and proficiency testing in cancer mutation analysis. Most of the existing controls for somatic mutations only harbor a single variant and are derived from unstable cell lines. This study aimed to establish a method to create stable multianalyte controls in a defined background by genome editing in GM12878 cells, which also can be applied for the reference of next-generation sequencing.

Methods: GM12878 cells were electroporated with a donor plasmid containing a mutant DNA sequence and a Cas9/sgRNA expressing vector. The genome-edited GM12878 cell was validated with Sanger sequencing, amplification refractory mutation system (ARMS), and next-generation sequencing (NGS).

Results: We have successfully generated a mutant GM12878 cell line harboring the defined variants including single-nucleotide variants (SNVs), small insertions and deletions (indels), and structural variants (SVs). The introduction of intended mutations in GM12878 cell line was confirmed by both ARMS and sequencing methods.

Conclusions: We developed a method for the preparation of the multiplexed controls for reference mutations in cancer gene by genome editing in GM12878 cells. This methodology can be used to generate other stable cancer reference materials with an unlimited supply.

KEYWORDS

control materials, CRISPR/Cas9, genome editing, next-generation sequencing, somatic mutation

1 | INTRODUCTION

Next-generation sequencing (NGS) technology has become a cornerstone in cancer management and has been widely adopted by clinical laboratories. Targeted NGS panels are commonly used to interrogate single-nucleotide variants (SNVs), insertions and deletions (indels), copy number variants (CNVs), and structural variants (SVs). Accurate mutation signature analysis is essential for guiding cancer therapy, definitive diagnosis, and predicting survival.

To ensure that NGS data are reliable and useful for clinical decision making, labs conducting NGS tests should comply with regulatory and professional quality guidelines.^{1,2} Quality control of NGS workflow is challenging because of its complexity. Control materials, which can be used to evaluate biases and errors in the NGS workflow, are necessary for test validation, quality control, and proficiency testing. The National Institute for Standards and Technology (NIST) has developed several reference normal cell lines (e.g., HapMap cell line GM12878) for whole-genome variant

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Clinical Laboratory Analysis* published by Wiley Periodicals LLC.

assessment.³ In the absence of reference cancer cell lines for whole genome sequencing, the normal samples are often used in the establishment of NGS test performance for somatic mutation analysis.⁴ The World Health Organization (WHO) also developed several gene-specific reference materials from cancer cell lines.⁵ Most reference materials for cancer mutation testing are derived from tumor cell lines which only contain a single mutation. To meet the demand for implementing multiplexed materials as controls for NGS assays, several approaches, including pooling characterized cancer cell lines and mixing synthetic DNA constructs, were developed. However, the cell line strategy may encounter the problem of sourcing, mixing, and characterizing. In addition, chromosomal instability and genetic drift, which affect the genomic mutation profile, are especially increased in cancer cells throughout cell culture. Synthetic DNA does not resemble patient's sample and cannot be used to monitor the extraction process.

Here, we aimed to develop a method to create stable and multiplexed control materials by genome editing in GM12878 cells, which also can be used for the control of matched tumor-normal sequencing. We previously generated genome-engineered cell lines containing intended mutations efficiently by harnessing clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated endonuclease Cas9 (Cas9)-induced nonhomologous end joining (NHEJ).⁶ Results showed that the engineered HEK293T and A549 cells harboring inserted synthetic mutant DNA fragments can serve as valid controls for genetic testing. In this study, a long mutant DNA fragment (covering SNVs, indels and SVs) for targeted NGS test was designed and introduced into the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) locus in GM12878 cells.

2 | MATERIALS AND METHODS

2.1 | Variant selection and donor plasmid

We selected five hotspot mutations (Table 1) in lung cancer that represent common variant types (SNVs, indels, and gene fusions). Each of the SNVs and indels was flanked by genomic sequence of *EGFR* exon and intron (300 bp both ends) region (Figure 1 and Table S1). For *EML4-ALK* variant 1 (E13:A20), a 1000-bp region both upstream and downstream of the breakpoint was designed for hybrid-capture NGS testing (Figure 1 and Table S1). A total of four small mutant DNA fragments were designed head to tail and assembled into a large fragment (Figure 1 and Table S1). In addition, two sgRNA

target sequences were added at both ends for cutting. The designed mutant DNA sequence (listed in Table S1) was *de novo* synthesized (Sangon Biotech) and cloned into pUC57 (GenScript) as a donor and further validated by sequencing.

2.2 | Cas9/sgRNA plasmid

SgRNA sequences targeting human *GAPDH* intron 7 were designed by using CHOPCHOP.⁷ Pairs of sgRNA oligos for each targeting site were cloned into pCAG-eCas9-GFP-U6-gRNA (Addgene # 79145) as previously reported.⁸ Target sequences of sgRNAs are listed in Table 2.

2.3 | Cell culture and electroporation

GM12878, a lymphoblastoid cell line, is a well-characterized Tier 1 cell of Encyclopedia of DNA Elements (ENCODE) project with a relatively normal karyotype. GM12878 was obtained from Coriell Institute and cultured in Roswell Park Memorial Institute 1640 medium (RPMI 1640; Gibco) supplemented with 15% fetal bovine serum (FBS; Gibco) and 1% penicillin-streptomycin (Gibco) at 37°C with 5% CO₂. The density of GM12878 cells was maintained at 3.5 × 10⁵ cells/ml before transfection. For each transfection, 0.6 pmol of Cas9/sgRNA plasmid, 0.6 pmol of donor plasmid, and 2,000,000 cells were prepared. The cells were washed twice with PBS and resuspended in 20 µl electroporation buffer (Celetrix), then the plasmids were mixed with the cells and transferred to a 20-µl electroporation tube (Celetrix). Electroporation was carried out at a voltage and pulse time of 440 V and 30 ms, respectively, using an electroporation machine (CTX-1500A LE, Celetrix). Following electrotransformation, the cells were transferred to a 24-well plate in warm medium.

2.4 | DNA extraction

Genomic DNA was extracted from cells using the QIAamp DNA Mini Kit (QIAGEN) according to the manufacturer's recommendations.

2.5 | T7 endonuclease I assay

Cells were collected 48 h after electroporation and genomic DNA was isolated. SgRNA targeted region (633 bp) was amplified

Plasmid	Mutations
A	NM_005228.3(<i>EGFR</i>):c.2369C>T (p. Thr790Met), NM_005228.3(<i>EGFR</i>):c.2573T>G (p. Leu858Arg), NM_005228.3(<i>EGFR</i>):c.2235_2249del15 (p. Glu746_Ala750del), NM_005228.3(<i>EGFR</i>):c.2310_2311insGGT (p. Asp770_Asn771insGly), <i>EML4-ALK</i> variant 1 (E13:A20)

TABLE 1 The donor plasmid harboring mutations

Abbreviations: *ALK*, anaplastic lymphoma kinase; *EGFR*, epidermal growth factor receptor; *EML4*, echinoderm microtubule-associated protein-like 4.

FIGURE 1 Schematics of creating multiplexed control materials for cancer mutation testing by integration of a mutant DNA sequence in GM12878 cells. Red arrow refers sgRNA cleavage site. ALK, anaplastic lymphoma kinase; *EML4*, echinoderm microtubule-associated protein-like 4; *GAPDH*, glyceraldehyde-3-phosphate dehydrogenase; NHEJ, nonhomologous end joining

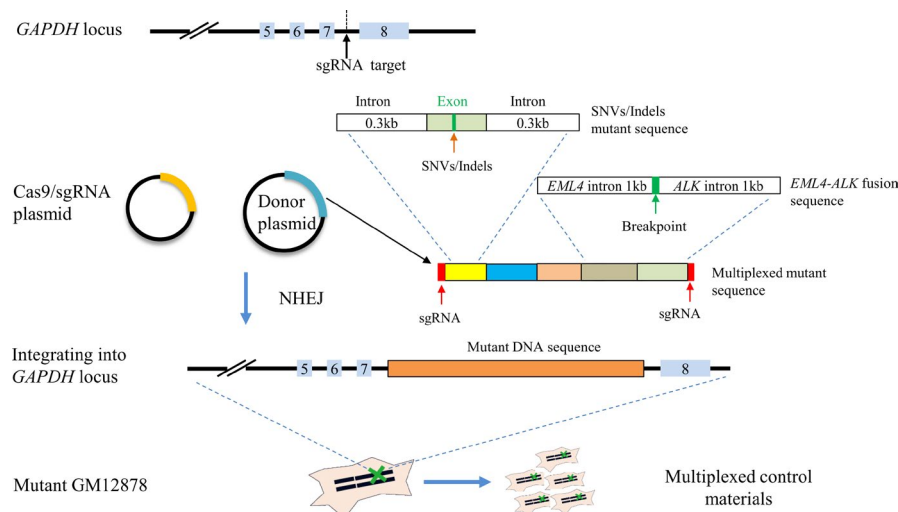


TABLE 2 Oligonucleotides used in this study

Name	Sequence (5'–3')	Intended use	Product length
SgRNA1	TCGATGGGTGGAGTCGCGTG	sgRNA target sequence	/
SgRNA2	GCGCAGGGTTAGTCACCGGC		/
SgRNA3	TAGCGTTGACCCGACCCCAA		/
T7E1-F	GGGGGACGCTTTCTTTCCTT	T7E1 assay	633 bp
T7E1-R	TTTCCGGAAGACGGAATGGG		
858F	ATGTCAAGATCACAGATTTTGGGCG	Allele specific PCR (L858R) for detection of integrated mutant DNA	128 bp
858R	CTGGTCCCTGGTGTCAGGAAAA		
egfr 19F	GCTGGTAACATCCACCCAGA	Sanger sequencing	/
egfr 19R	GAGAAAAGGTGGGCTGAG	Sanger sequencing	/
egfr 20F	CCTCCTTCTGGCCACCATGCG	Sanger sequencing	/
egfr 20R	CATGTGAGGATCCTGGCTCC	Sanger sequencing	/
egfr 21F	CGGATGCAGAGCTTCTCCC	Sanger sequencing	/
egfr 21R	AGGCAGCCTGGTCCCTGGTG	Sanger sequencing	/

Abbreviations: ALK, anaplastic lymphoma kinase; *EML4*, echinoderm microtubule-associated protein-like 4; T7E1, T7 Endonuclease I.

using high-fidelity polymerase chain reaction (PCR) master mix (ThermoFisher) and purified with zymoclean Gel DNA Recovery Kit (Zymo). Then, 200 ng of purified PCR products were denatured, annealed, and incubated with T7 Endonuclease I (T7E1) (NEB) following manufacturer's instructions. The cleavage efficiency of sgRNA was analyzed by using 2% agarose gel. PCR primers are listed in Table 2.

2.6 | Fluorescence-activated cell sorting

Cells were harvested 48 h after transfection and resuspended in PBS containing 0.5% bovine serum albumin (BSA) and 2% penicillin-streptomycin for sorting. Single GFP-positive cells were sorted in 96-well plates using MoFlo Astrios EQ (Beckman Coulter). Cells were cultured in 50% conditioned media for 4–5 weeks and then transferred to 24-well plates upon expansion.

2.7 | PCR screening of positive clones

Allele-specific PCR (detecting mutation L858R) was carried out for detection of integration of mutant DNA sequences in cell clones. DNA extracted from single clones was amplified using GoTaq Green Master Mix (Promega) and primers (listed in Table 2) following the manufacturer's instructions.

2.8 | Sanger sequencing

The designed variants in the positive cell clones were validated by Sanger sequencing. Briefly, genomic DNA was first amplified with specific primers⁹ (listed in Table 2), then the PCR products were purified, and sequenced by using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on the ABI 3500DX Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions.

2.9 | Amplification refractory mutation system

The designed *EGFR* variants in positive clones were tested using amplification refractory mutation system (ARMS) with a certified *EGFR* mutations detection kit (AmoyDx) according to the manufacturer's instructions. Tests were performed on ABI-7500 real-time PCR system.

2.10 | Next-generation sequencing

Positive cells were further tested by targeted NGS. Essential NGS panel (AmoyDx), an approved hybridization capture-based NGS kit (hotspot cancer panel), was used. Tests were conducted on Illumina Nextseq 500 system (Illumina) following the manufacturer's instructions.

3 | RESULTS

3.1 | Plasmid construction

We designed 3 sgRNAs to target the 7th intron of *GAPDH* gene. The cutting effectiveness of each sgRNA was assessed by the T7E1 assay using the genomic DNA extracted from the electroporated cells. Based on results (Figure S1) of the T7E1 assay, we selected sgRNA2 for further experiments. The donor plasmid harboring the corresponding mutant DNA sequence (Table 1) was constructed and validated via Sanger sequencing.

3.2 | Generation of modified GM12878 harboring the intended mutations

Following transfection, we plated 960 individual cells after sorting. Most of the cells cannot survive electrotransfection and single cell culture. Only one positive clone (clone 24) was obtained among the 60 expanded and screened clones, which was designated as 12878-Multiv. Figure 2 showed the representative PCR screening results. We further validated target variants in this positive clone by using ARMS method and Sanger sequencing (Figure 3 and Figure 4).

3.3 | Variant detection using NGS

Genomic DNA NGS library was captured for target enrichment and sequenced on the Nextseq 500 system (Illumina). The sequencing data were analyzed according to the manufacturer's instructions. The mutations in the genome-edited cell were successfully detected (Figure 5).

4 | DISCUSSION

In this study, we created an engineered cell line from GM12878 using CRISPR/Cas9 technology. This cell line can be used as a multiplexed mutant standard for various detection methods including targeted gene sequencing. Paired with GM12878, they also can be used as controls for tumor-normal sequencing. Because of its high transfection efficiency, HEK 293 are frequently used to produce genome-edited cell lines as controls for somatic mutation testing.^{6,10-12} However, HEK293 is a hypotriploid cell line which displays some cytogenetic instability and is not well-characterized across the genome. On the contrary, GM12878 is stable, and most importantly, it is extensively characterized and serves as benchmark reference materials for NGS testing. The use of electroporation in our study offered transfection efficiencies up to 40% in GM12878, which is acceptable for downstream experiments. One shortcoming of using GM12878 is that it is difficult to grow into single clones after electroporation and FACS. Feed cells may be used. Generation of mutant cells by mutation knock-in via on homology-directed repair is another approach, however, the efficiency is relatively low.¹² In addition, only nucleotides adjacent to protospacer adjacent motif (PAM) site can be efficiently edited with this strategy, which may exclude many clinical actionable variants. By using CRISPR/Cas9-induced homology-independent integration, we can efficiently introduce any long DNA sequence harboring design variants into any locus of chromosome. We did not evaluate the off-target effects because the artificial mutations were generated by inserting long DNA sequences harboring variants to the chromosome locus of an unrelated gene. In such a way, neither the indels in the integration site nor the insertion in the off-target sites will have effects on the detection of the designed variants. Nonetheless, the engineered high-fidelity eCas9¹³ was used in our study to minimize off-target effects. Furthermore, the

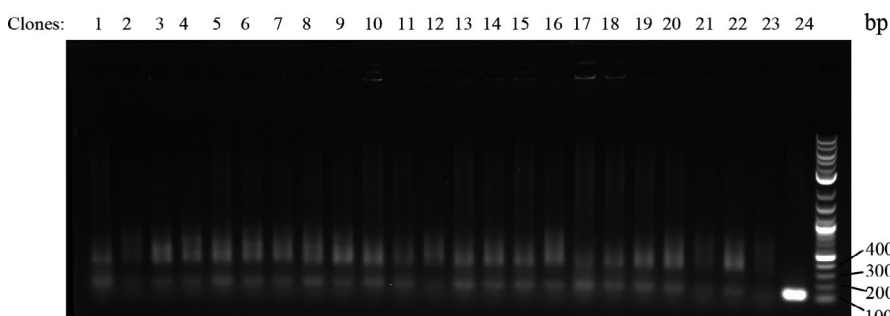


FIGURE 2 PCR screening of the positive clones in genome-engineered GM12878 cells

FIGURE 3 Validation of the designed variants in genome-edited GM12878 cell by Sanger sequencing. *EGFR*, epidermal growth factor receptor

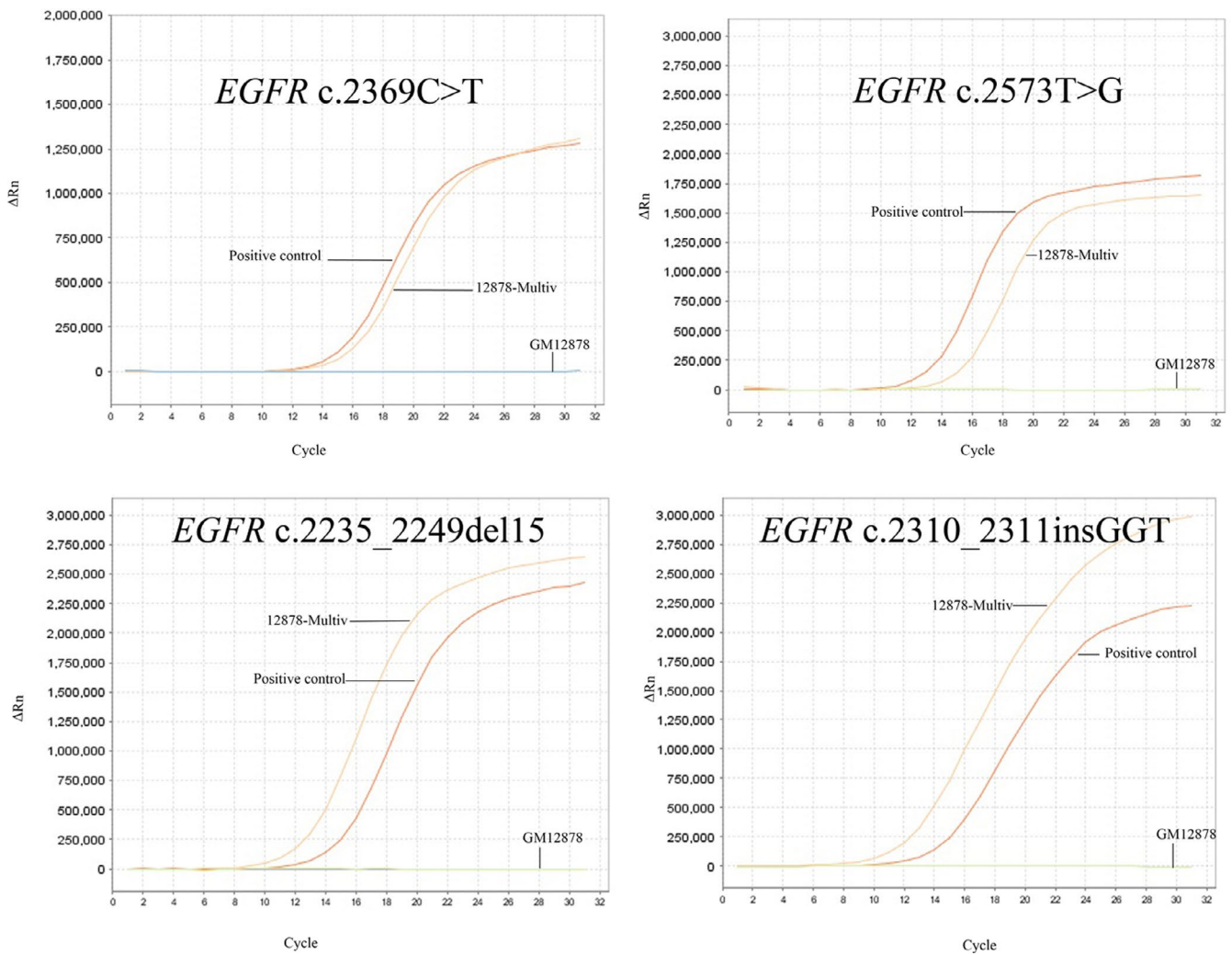
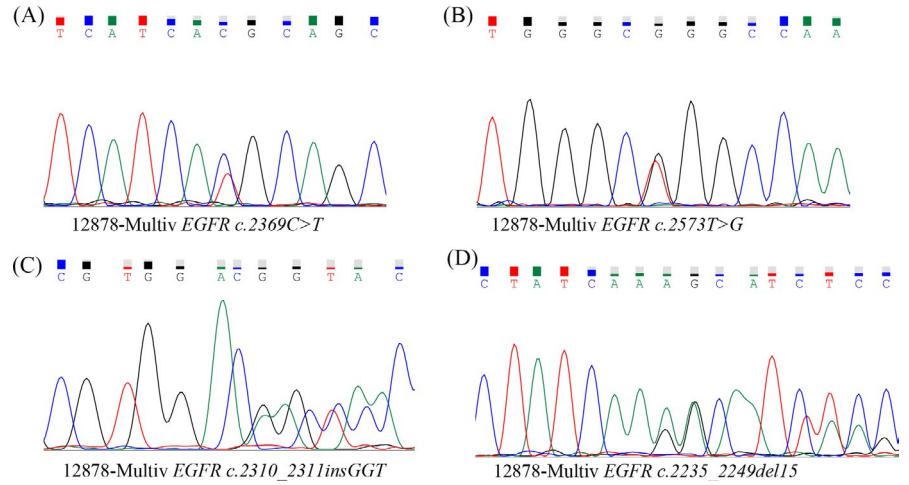


FIGURE 4 Verification of the intended variants in genome-edited GM12878 cell (12878-Multiv) by ARMS. *EGFR*, epidermal growth factor receptor

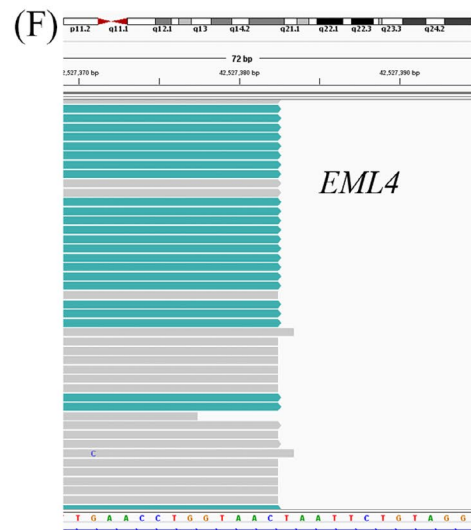
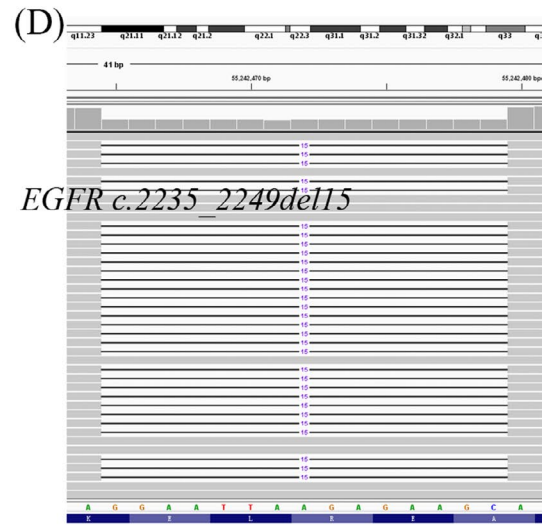
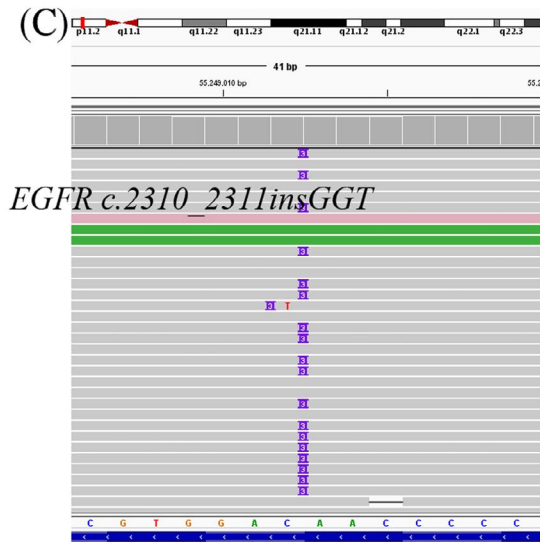
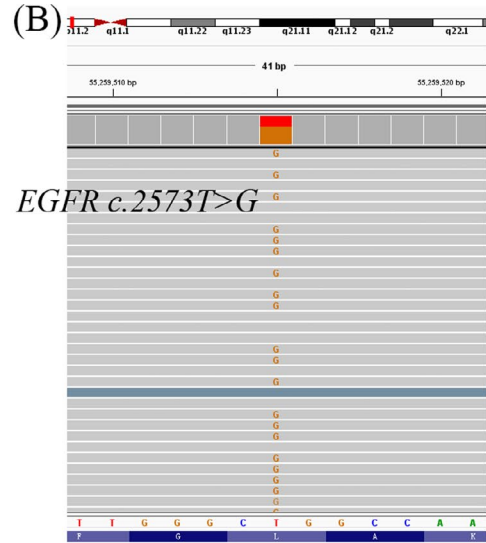
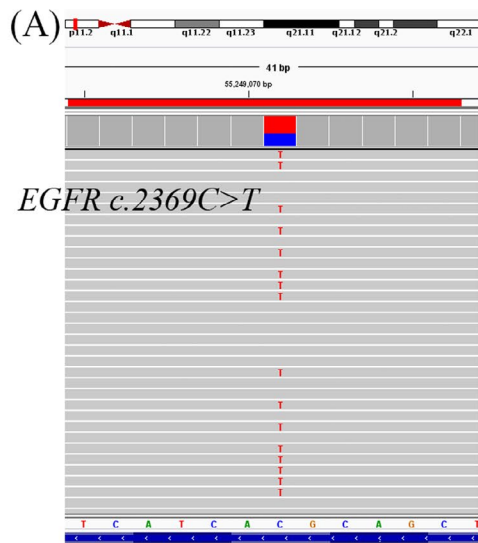


FIGURE 5 Integrative Genomics Viewer (IGV) screenshots of the target variants in genome-edited GM12878 cell (12878-Multiv) detected by NGS. (A) variant allele fraction (VAF) = 56.9%; (B) VAF = 64.3%; (C) VAF = 54.9%; (D) VAF = 50.8%; (E and F) VAF = 33.4%; ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor; EML4, echinoderm microtubule-associated protein-like 4

methodology developed in this study can be extended to create a wide range of cancer reference materials.

The results reported herein should be considered in light of some limitations. Since this is a pilot study, only five hotspot mutations were included in the engineered cell line. In addition, our developed control containing the artificial EML4-ALK mutation cannot be used for RNA sequencing. Follow-up studies are necessary to cover more variants in cancer panel. Finally, these control materials also need to be tested across different NGS platforms to establish broad utility.

ACKNOWLEDGEMENTS

This work was supported by the Beijing Municipal Natural Science Foundation [7192188, 7204299], Beijing Dongcheng District Outstanding Talent Nurturing Program [2019WJGW-10-01], and National Natural Science Foundation of China [81974319, 81902145].

DATA AVAILABILITY STATEMENT

The sequencing datasets generated during the current study are available from the corresponding author on reasonable request.

ORCID

Guigao Lin  <https://orcid.org/0000-0002-0406-9078>

Jinming Li  <https://orcid.org/0000-0002-1476-3397>

REFERENCES

- Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol.* 2012;30:1033-1036.
- Aziz N, Zhao Q, Bry L, et al. College of American pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Lab Med.* 2015;139:481-493.
- Zook JM, Catoe D, McDaniel J, et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data.* 2016;3:160025.
- Jennings LJ, Arcila ME, Corless C, et al. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the association for molecular pathology and college of American pathologists. *J Mol Diagn.* 2017;19:341-365.
- Sanzone P, Hawkins R, Atkinson E, et al. *Collaborative Study to Evaluate the Proposed who 1st International Reference Panel for Genomic Kras Codons 12 and 13 Mutations.* World Health Organization; 2017.
- Lin G, Zhang K, Peng R, Han Y, Xie J, Li J. Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-associated endonuclease cas9-mediated homology-independent integration for generating quality control materials for clinical molecular genetic testing. *J Mol Diagn.* 2018;20:373-380.
- Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 2019;47:W171-W174.
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* 2013;8:2281-2308.
- Xie G, Xie F, Wu P, et al. The mutation rates of EGFR in non-small cell lung cancer and KRAS in colorectal cancer of Chinese patients as detected by pyrosequencing using a novel dispensation order. *J Exp Clin Cancer Res.* 2015;34:63.
- Peng R, Zhang R, Lin G, et al. CRISPR/Cas9 technology-based xenograft tumors as candidate reference materials for multiple EML4-ALK rearrangements testing. *J Mol Diagn.* 2017;19:766-775.
- Jia S, Zhang R, Lin G, et al. A novel cell line generated using the CRISPR/Cas9 technology as universal quality control material for KRAS G12V mutation testing. *J Clin Lab Anal.* 2018;32:e22391.
- Suzuki T, Tsukumo Y, Furihata C, Naito M, Kohara A. Preparation of the standard cell lines for reference mutations in cancer gene-panels by genome editing in HEK 293 T/17 cells. *Genes Environ.* 2020;42:8.
- Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science.* 2016;351:84-88.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lin G, Zhang K, Han Y, Peng R, Li J. Preparation of multiplexed control materials for cancer mutation analysis by genome editing in GM12878 cells. *J Clin Lab Anal.* 2022;36:e24139. doi:[10.1002/jcla.24139](https://doi.org/10.1002/jcla.24139)